ASSESSING OCEAN MODELING AND DATA ASSIMILATION REQUIREMENTS

An activity sponsored by the National Science Foundation,
Ocean Sciences Division

1. Report on Workshop to Discuss Needs for an Ocean Data Assimilation Center

Adam's Mark Hotel, Dallas TX
20-21 February 1997

May 1997

This document may be cited as:

Report on Workshop to Discuss Needs for an Ocean Data Assimilation Center, Dallas, February 1997, 18 pp., U.S. WOCE Office, College Station, Texas, 1997

Copies of this report may be obtained from the U.S. WOCE Office, Department of Oceanography, Texas A&M University, College Station, TX 77843-3146, (409) 845-1443, uswoce@ocean.tamu.edu

TABLE OF CONTENTS

REPORT ON WORKSHOP TO DISCUSS NEEDS FOR AN OCEAN DATA ASSIMILATION CENTER

Adam's Mark Hotel, Dallas TX
20-21 February 1997

I. BACKGROUND

In August 1996, the U.S WOCE Science Steering Committee held a meeting in Washington, D.C., to discuss the synthesis phase of WOCE and, in particular, whether WOCE plans for synthesis fit the ongoing needs of U.S. federal funding agencies. Representatives from DOE, ONR, NASA, NOAA and NSF were present. It was soon realized that most agency operational needs are for seasonal-to-interannual predictive capability, rather than the decadal-scale variability studied by WOCE. It was also realized, however, both that WOCE data can provide information important on shorter scales and that the problem of data assimilation is one that will shortly be faced by many other global programs within the U.S. GCRP, and perhaps by many individual researchers as well.

Some individuals and groups as part of large research programs are already using ocean data assimilation for various purposes. Data are assimilated into ocean regional, basin, and global models for hindcasting (e.g., WOCE at basin and global scales for synthesis), nowcasting (currents and water levels for major U.S. harbors and ports in operational mode by NOAA), and forecasting (operational mode for ENSO prediction).

Some of the large U.S. ocean research programs either now use or plan to use ocean data assimilation for various research purposes, e.g., data synthesis or research into improved nowcasting and forecasting. The elements most advanced in this type of work are coastal and shelf oceanography, GOALS (ENSO interests), and WOCE. They are working on different problems and are at different levels of expertise in attacking their problems: for example, ENSO modelers have considerable experience, there are a variety of coastal models that assimilate data, but only very few experimental basin or global scale test syntheses have been carried out using full ocean models. Some elements of the ocean science community have not yet seriously begun to combine data with models. The points are that (1) there is a need for additional research to learn how to effectively use data with models for oceanographic synthesis, and prediction when feasible and (2) different programs and segments of the community are at different levels of learning regarding how to do this.

Over a two-year period, the U.S. WOCE community attempted to identify existing federal laboratories where concerted, long-term data synthesis might be supported. The response was that existing laboratories have mandates that preclude their use for community development of ocean data assimilation capabilities. Of course, this might change if additional, new funds were identified. Another option is to develop new capabilities to meet the needs. In either case, considerable funding might be needed to sustain long-term efforts to develop community capabilities in ocean data assimilation. Support for such activities is not available now within the present funding envelopes.

Having knowledge of the background and an appreciation that current projections are for essentially level funding in the near term, the Ocean Sciences Division of the National Science Foundation, directed by Mike Purdy, is considering future requirements for infrastructure needed by the ocean sciences community in the pursuit of ocean modeling, including data assimilation. Purdy send a message on August, 30, 1996 to Worth Nowlin stating:

> "...we believe that an important component of the future global ocean sciences is the creation of the infrastructure and environment in which data assimilation, integration, modeling, and interpretation of large diverse data sets can take place. This is an issue that extends well beyond WOCE and we encourage you to lead a community-based effort to provide us with advice concerning the form that support for such capability should take. We consider it important to design a model that can sustain growth over a substantial period of time because it is inevitable that the beginnings will be modest, but the requirement will grow substantially over the next decade.
>
> It is logical that WOCE should provide the leadership for this activity, but it is important that the effort be...a service to the community as a whole. ... Would it be useful, perhaps, to form a small high-level 'executive' steering group...to serve as a leadership group and attempt from the start to establish the broad non-partisan nature of this initiative?
>
> In recognition of the broad-based nature of this planning activity we...would supplement it at the Division level...."

An ad hoc steering committee for ocean modeling advances consisting of Andrew Bennett (community at large), Russ Davis (ocean CLIVAR), Hugh Ducklow (JGOFS), Worth Nowlin (WOCE), Thomas Powell (GLOBEC), and Doug Wallace (DOE carbon dioxide program) was established in September 1996. A series of community discussions to assess requirements for future computational work by the ocean science community was considered.

The first meeting in this series concentrated on synthesis needs. The intent of the meeting was to begin the process of exploring community requirements for data assimilation and to discuss whether a community ocean data assimilation center is required.

Invitations were sent to representatives of CLIVAR, CoOP, DOE Carbon dioxide program, GLOBEC, JGOFS, ODP, RIDGE, and WOCE, to representatives of NASA and NOAA laboratories, and to several individual scientists in the community who could contribute expertise (Unfortunately, not all who were invited could attend.) The meeting was open and an announcement was published in WOCE Notes. Names of those who attended and their affiliations are given in Appendix I.

Background papers, dealing particularly with requirements of NASA and WOCE, and a suggestion from the Ocean Observations Panel for Climate of the Global Ocean Observing System/Global Climate Observing System, were circulated to participants prior to the meeting. Other invitees were given the opportunity to prepare background papers if they so wished. These also were distributed to attendees and other interested parties prior to the meeting.

A second meeting, organized by a committee consisting of Dale Haidvogel, Jim McWilliams, Ken Denman, and Rainer Bleck will be held in Boulder, CO in April 1997. The focus will be on community needs in support of ocean general circulation modeling. In addition the NRC committee to consider coordination and needs of large ocean science programs, chaired by Rana Fine, will address the issue of future computational needs at a meeting during this summer; that forum may provide additional information.

II. STRUCTURE OF MEETING

The agenda for the meeting is given in Appendix II. Following a resume of the background to the meeting by Nowlin, Dick Lambert and Mike Reeve reminded the group that oceanography is presently at a crossroads. Field programs of many large programs are winding down or will be in the not so distant future, and the best feasible approaches to synthesis remain somewhat unclear. Moreover, other programs are waiting to begin. Thus, resources for data synthesis will, necessarily, be limited; there is the need to identify what resources are essential and to improve access to them. It is likely that many currently missing resources are common to several or all programs, and thus it is important to ensure that the entire community be involved in laying out requirements.

The meeting began with brief introductory statements from attendees on their interests/requirements (or those of the program they were representing) and aspirations for the meeting. That proved a very helpful and informative manner in which to begin the meeting because it tabled most of the needs and questions.

There followed a series of four presentations by Andrew Bennett (General elements of ocean data assimilation), Keith Thompson (Shelf models), Dave Behringer (Quasi-operational ENSO prediction modeling) and Jochem Marotzke (Basin/global ocean modeling). Resumes of these talks are given in Appendix III.

Time was allotted to representatives of large research programs or laboratories to make fuller statements and introduce specific plans and needs for data assimilation by that program or laboratory. These were discussed by the meeting participants.

The second day of the meeting was devoted to discussion, focused on several specific themes with discussion leaders:

(1) Common requirements
(2) Special requirements for biogeochemical studies
(3) Components of ocean data assimilation needs (infrastructure)
(4) Sociological factors to be considered
(5) Technical aspects of data assimilation

Most discussion of the last item was scattered throughout the meeting.

III. PLANS AND REQUIREMENTS OF SPECIFIC PROGRAMS AND LABORATORIES

CLIVAR

The TOGA program has shown that predicting seasonal-to-interannual variability in the ocean is possible. WOCE is concentrating on collecting data which may be used for a similar purpose for longer-term predictions. TOGA ended in 1994 and WOCE will be completed within the next few years. A follow-up program, CLIVAR, has been formulated to increase our understanding of, and ability to predict, climate variations on time scales of seasons to centuries. At present, CLIVAR has no published implementation plan, but this is being developed by several groups. As a result, no specific requirements could be stated although it is likely that the program will have requirements very similar to the other global programs. It was also pointed out that there is a worrying tendency for agencies to support forecasting groups (particularly for seasonal-to-interannual prediction) while simultaneously cutting support for the operating systems that supply the data needed to initialize the models used for the forecasting.

CoOP
This program seeks to obtain a new level of quantitative understanding of the processes that dominate the transports, transformations and fates of biologically, chemically and geologically important material on continental margins. This includes particularly cross-margin transport. The program is operating through a series of process studies, coupled with modeling. Data assimilation may therefore be an important aspect of CoOP synthesis.

Typical applications could include: studying what parameters link atmospheric forcing with variations in cross-shelf flow and sediment and larval transport; how gases and other materials are transported between the coastal atmosphere and ocean; what processes control the cross-shelf transport of materials in the Great Lakes; and the development of coupled physical/biological models in which physical and biological data are assimilated into predictive schemes. There is also need for continued support for coastal time-series sites.

GLOBEC
The GLOBEC program has the objective of understanding ocean ecosystem dynamics and how they are influenced by physical processes so that the predictability of population fluctuations in a changing global climate can be assessed. The present focus is on the north-west Atlantic, although a major program in the north-east Pacific is due to commence shortly. The program concentrates mainly on the near-shore (< 50 miles) zone, but can extend up to about 500 miles offshore.

The main requirements at present are for improvements in linking physical and biological models and embedding regional models within basin- and global-scale models. This will require different governing equations from those ordinarily used in physical oceanography. Additionally, the intensive computation needed for the above means that scientists are tending to approach the operating limits of present computers.

JGOFS
The primary objective of U.S. JGOFS is to improve our understanding of the processes controlling the biogeochemical fluxes of carbon within the ocean and across the air-sea and ocean-sediment interfaces. The field program has three components: a global survey of the large-scale inorganic carbon dioxide system (in conjunction with WOCE); two time-series stations off Bermuda and Hawaii which sample approximately monthly; and a series of intensive process studies lasting up to one year. Synthesis and modeling of the data sets will continue through 2001 or 2002.

JGOFS sees two potential uses for any ocean data assimilation center. The first is to develop improved physical models for interpreting and synthesizing the various data sets (e.g., to allow global inversions for estimating dissolved inorganic carbon transport or the air-sea flux of carbon dioxide, to support retrospective studies of the time-series and process study data, and to permit work on more detailed mesoscale/regional problems). The second is to assist with improvements to biogeochemical models (e.g., by developing new equations for parameter estimation, improving the extrapolations from regional to global scale, and developing methods for using satellite-derived ocean color data).

On the longer term, it is likely that data assimilation will be crucial for addressing important questions relating ocean biogeochemistry to climate change. An ocean data assimilation center could serve as one venue for the exchange of information between physical, chemical and biological oceanographers.

ODP

The Ocean Drilling Program is a major component of the general program in paleo-oceanography, which is examining changes in past climates via a series of proxy data. The time scales of interest cover very wide ranges - from seasonal changes (such as shown by corals) through decadal (e.g., rapidly accumulating sediments) to very long term (centuries to millennia - ODP).

Main requirements for data assimilation are methods by which sparse data, generally representing time-averaged climatologies, can be integrated into three-dimensional models. There is the need to construct physically meaningful fields which reflect model simulations of past conditions. Because the data represent past ocean conditions, it is frequently unclear what boundary conditions or other parameterizations should be used. Also, it is equally unclear whether the proxy paleo data should be used as input data, model constraints, or otherwise.

As an example, modern SST/phytoplankton assemblage data are used to determine regression equations linking plankton distributions to surface temperature. These same regression data are then used with paleo distribution data to infer past temperature distributions. The CLIMAP program showed that during Ice Ages, while polar plankton distributions expand equatorwards, there is little change in the distribution of tropical species. Thus the subpolar and subtropical groups are squeezed into narrower latitude bands.

RIDGE

Given the assumption that the global spreading-center network may be viewed as a single, dynamic system of focused energy flow from the earth's interior to the lithosphere, hydrosphere, and biosphere, the primary goal of the RIDGE initiative is to understand the geophysical, geochemical, and geobiological causes and consequences of that energy transfer within the global rift system through time.

Some of the questions that are of interest to the RIDGE community that might be addressed using an Ocean Data Assimilation Facility are as follows:

1) What is the effect of hydrothermal inputs on ocean circulation? Do these inputs (either steady-state hydrothermal inputs or episodic inputs, such as megaplumes) drive or modify ocean circulation?

2) How is the hydrothermal effluent dispersed? (The RIDGE community has both physical and chemical data from hydrothermal plumes at numerous sites throughout the world ocean.)

3) How are the larvae associated with vent organisms dispersed? (There is a project within RIDGE–LARVAE–that focuses on this question. Its field component will begin within the next year.)

4) What is the effect of bathymetry on ocean circulation? Can we predict circulation near ridges; what is the effect of other features such as transforms or fracture zones on the circulation; and what is the effect of other topographic features, e.g., islands, on circulation? (The RIDGE community has collected an extensive amount of bathymetric data.)

5) What is the circulation pattern of hydrothermal fluids within the oceanic crust?
- on axis, high temperature (focused flow) hydrothermal systems
- on axis, low temperature (diffuse flow) hydrothermal systems
- off axis, low temperature hydrothermal systems within both the crust and sediment cover

6) How do geodynamic models of mantle flow (physical properties) and petrologic models of melt generation (chemical tracers) compare with seismic, geochemical, gravity and bathymetric data? Can these be combined in a joint inversion of geophysical and geochemical data? (Although the physical properties are different, this is analogous to the problem of combining physical with chemical oceanographic data sets—as WOCE is doing.)

7) Are there temporal relationships between seismicity, hydrothermal venting, and biological productivity? (i.e., Time series studies linking large and different types of data sets.)

8) Can global models of mantle flow and plate motion be combined with bathymetric and
petrologic data to see if they can reproduce regional trends in basin scale bathymetry and long wavelength variations in axial depth and chemistry?

9) An important issue for the 'Event Detection & Response' part of RIDGE is better understanding T-phase signals, how they are formed, how they propagate, how they are affected by the physical oceanographic properties of the water column? (At present this is related to the SOSUS detection of magmatic events in the northeast Pacific. The power of this technique to address issues of crustal accretion will likely lead to moorings in other parts of the world ocean as well.)

Much of the RIDGE community is not conversant with using integrated computing facilities and large computer models to handle either large data sets or data sets of multiple types. The RIDGE community is by definition also very multidisciplinary. Hence for an Ocean Data Assimilation Center to be an important asset to the RIDGE community requires it to have the infrastructure available to help RIDGE researchers to work jointly with computational scientists to learn what tools are available to approach interdisciplinary problems, and how best to apply them to RIDGE-related problems.

WOCE
The WOCE program has the twin goals of developing models useful for predicting climate change and collecting the data necessary to test them, and of determining the representativeness of the specific WOCE data sets for the long-term behavior of the ocean, in order to find methods for determining long-term changes in the ocean circulation. On the practical side, there is also the aim of identifying, and hopefully initiating, continued long-term measurements that can monitor the variability of the ocean climate. Thus WOCE results feed directly into the CLIVAR Dec-Cen program, as well as providing the global physical oceanography framework for programs such as JGOFS and GLOBEC.

Thus the main requirements for data assimilation within WOCE concern the provision of a dynamical framework for analysis of the observations, the placing of the observations within their historical context, and the validation of the surface boundary conditions and other parameterizations used in global circulation models. A broad range of synthetical and assimilation activities is envisaged, as detailed in two recent documents (U.S. WOCE Office 1996 and IPO 1997). In particular, there exists the need to assimilate within one model data from diverse platforms such as temperature and salinity data from PALACE floats, hydrographic surveys and VOS observations; direct current data from floats, drifters and current meter arrays; sea level data from coastal stations and satellite altimetry; and surface forcing fields from both in situ observations and satellites.

NASA
The latest generation of satellites are providing near-global coverage of sea surface height (from altimetry), sea surface temperature, and wind fields (from scatterometry). These measurements are all essential for global ocean models, and the data are being assimilated into such models with more or less success. Sea-ice and ocean color data either are, or soon will be, available as well. NASA's requirements for ocean data assimilation were discussed at a meeting in Irvine in 1994 (NASA, 1996). The meeting recommended the development of 4-D assimilation capabilities through group activities at designated "applications centers" that act as bridges between individual researchers at universities and operational centers such as National Center for Environmental Prediction. Additionally recommended was a coordinated national program on assimilation, funded by all interested federal agencies, to optimize use of resources. While NASA's own mission is concentrating on seasonal-to-interannual variability, there is a move to establish a longer-term capability in cooperation with NSF.

NOAA
This agency supports several ongoing data assimilation efforts, all of which are aimed at resolution of the ENSO problem and seasonal-to-interannual variability. At NCEP, the purpose of data assimilation is to provide a near-real-time analysis of the state of the ocean and to initialize a coupled ocean-atmosphere model for seasonal forecasts. It has been found that the performance of the ocean model used is critical to both these operations, and there is need for better understanding of how the model variability corresponds to reality. Additionally, use of the coupled model may require an assimilation system that allows filtering of the ocean initialization to suppress error growth, or a system that allows initialization of only certain empirical modes of ocean variability.

GFDL
This laboratory also has plans to develop ocean data assimilation at the Center for Ocean Data Assimilation and Modeling (CODAM). The center will concentrate on predictions in the following three areas: coastal and mesoscale variability, seasonal-to-ENSO prediction, and decadal prediction of the North Atlantic thermohaline circulation. It is anticipated that the work will cover both OGCMs and coupled models. Data assimilation will be used both to improve models (through the use of adjoint methods) and to produce products. Examples of likely products include model-enhanced climatologies and meridional fluxes for the North Atlantic.

NCAR
The new (1994) initiative, the Climate System Model (CSM), has the long-term goal of building, maintaining, and continually improving a comprehensive model of the climate system, including both physical and biogeochemical aspects. Although this is a forward modeling operation, several opportunities exist for collaboration with any ocean data assimilation center once it is developed. The CSM can contribute new forward modeling technologies, such as new parameterizations to reduce bias. Additionally, the CSM can use outputs from an ocean data assimilation center for initialization of their own models, particularly for seasonal-to-interannual time scales, and could link with the assimilation to evaluate, calibrate and optimize elements of the forward models.

IV. IDENTIFICATION OF COMMON REQUIREMENTS

Although the participants at the meeting represented widely-varied groups of researchers with very differing time and space scales of interest, there was considerable commonality between them. It seems important to stress the broad dynamic range of modeling activities that are common across the ocean sciences community. Examples include common needs for (a) eddy-resolving and non-eddy-resolving models; (b) detailed and elaborate models versus less detailed, simple models; and (c) development of model physics and codes for difficult, unresolved processes, e.g., upper layer processes and biogeochemical processes.

Common requirements were identified by the meeting as being important to future data assimilation efforts. These are arranged into groups related to data, models, sociological aspects, and users and products.

Data

1. <u>Continuing data</u> There is the need for data gathering operations to continue. This may seem obvious, but given the perceived tendency for agencies to cut support for operational data-gathering systems at the same time as they are expanding support for modeling and synthesis activities, it requires stating. Without such continued support the data flow required by the modeling community will cease. In-situ data are particularly important, as they can be used to complement and validate remotely-sensed data such as are being provided by satellite systems.

2. Timely data access   A data synthesis center will need rapid access to certain datasets, depending on the type of model being used. Timely access to such data is a prerequisite of a successful data assimilation effort.

3. Estimates of data error and scales Data assimilation requires more than just data sets. These need to be organized and sampled in such a way that data assimilators can obtain correlation forms and scales for each variable studied. Among others, requirements exist for estimates of instrumental, sampling and other noise (errors). These allow modelers to make assessments of the uncertainty of results obtained. In addition, such assessments will provide information to improve planning of components of the long-term observing systems and motivate new fieldwork.

4. Retrospective analyses Needed are continuing analyses of data sets to provide historical perspective with error bars. Long time series with known levels of quality are essential in detecting and describing variability.

Models

5. Continuing model development There is a need for continued model development, including the development of new codes, priors for each model (see Bennett's talk in Appendix III) and specific mathematical investigations related to these. Improved model physics is understood to be a goal of data assimilation. Also, new methods of doing data/model comparisons must be investigated. For example, some parts of a data set, or a complete set, can be withheld from an assimilation to see whether the model can reproduce the withheld data.

6. Community models There is a need to support several models as community models, although no attempt was made to define how many are required.

7. Assess computational requirements There is a need to assess the computational requirements for present and future operations in this area. Present models are becoming limited by the availability of hardware. This is likely to become more acute as researchers try to insert geochemical and biological models within physical models and improve resolution.

Users and products

8. Users Strong focus should be placed on users of model products. Initial users likely will be researchers. However, it is important to plan from the beginnings of organized ocean data assimilation for the accommodation of a broader range of users. These could range from state and federal agencies with mandated requirements for data products, to special requirements of the private sector, to needs of policy makers for assessments. The presentation by Keith Thompson regarding uses of products from operational shelf/coastal models in Canada (see Appendix E) was especially illustrative of the potential for accommodating users with synthesis products.

9. Products Consider the impacts of COADS, Levitus climatologies, and similar products on research modeling. Clearly such analyzed fields will be primary products of data assimilation. There is the need for each of the large programs and disciplinary areas to identify needed products. One such product is flow fields—these are required by ocean scientists in most study areas. There is also the need to ask what products can be produced in the course of research that might be of value to non-research elements of society.

Sociological aspects

10. Education Teach people, by example, what to expect of data assimilation. Objective interpolation provides a framework. But, most people don't understand the limitations/ drawbacks of analysis by statistical interpolation. It will be a large next step to interpret the evolution of sets of maps (representing states of the ocean) to better understand why such changes occur.

11. Improved leadership There is a need to improve elements of leadership for new efforts such as an ocean data assimilation center. It is important to ensure that established objectives are reasonable. Community leadership must be joint between agencies and academia.

12. Connectivity The community must be encouraged to establish and maintain active connections between modeling groups within diverse programs and disciplines. An ocean data assimilation center must maintain connections with other centers, data sets, and related activities. Improvements in the connectivity between modelers, code writers, and observationalists is a requirement.

13. Design of observations Ocean data assimilation has a natural role in planning/design for new field programs and long-term observing systems. This requires use of model output by observationalists, and thus mutual interest.

Summary

It was generally agreed that the ocean science community would greatly benefit from the capabilities available through an ocean data assimilation center. The evolution of states of the ocean could be simulated. Forecasts would remain the longer-term goal. The recommendation was that the community work to establish an ocean data assimilation center. However, no consensus was reached concerning whether such a center should be centralized or distributed.

Specific requirements include access to continuing data streams, with continued support for ongoing model and code improvements. Critical levels of support are necessary for the center. The main aim of such a center will be to prepare/archive/distribute useful products; examples of these include analyzed fields of scalar properties (e.g., salinity and temperature) and improved flow fields. Additionally, the center will develop the capability to marry models of shelf/coastal regimes with the open ocean and incorporate biological/ biogeochemical models with physical models.

V. INFRASTRUCTURE REQUIREMENTS OF AN OCEAN DATA ASSIMILATION CENTER

The meeting discussed desirable characteristics of an ocean data assimilation center (a national ocean synthesis activity). This section assumes that a physical center will be established. It was agreed that this activity must:

- Emphasize the need for (super)critical levels of effort;
- Adopt an interdisciplinary outlook; and
- Focus on products.

Elements of the activity must include:

- Ready access to multiple community datasets, products, and climatologies (it is not necessary for the center to be situated at a major data center but there must be good networking links between them – presently datasets of up to several Gbytes can be handled);
- Capability to run several community models at different resolutions (e.g., global, coastal and biogeochemical models);
- Ready access to the latest technology for data assimilation;
- Critical scientific and support mass;
- Centralized computing necessary for long-term computations; and
- Strong outreach and visitor programs.

In discussing the goals for such a center, it was agreed that the probable first step should be production of products based on of one or more physical model(s) of the global ocean at a particular resolution. It is anticipated that successful physical models will evolve to incorporate biogeochemical and biological models, as well as being seen as a framework for supporting nested coastal models. This will require additional funding support. The center must have funds to host visits by scientists from the modeling, observational, and data assimilation communities.

No conclusions were reached on whether there should be only one such center, but the recommendation was that at least one is necessary.

VI. NEXT STEPS

The meeting anticipated several particular follow-on activities related to the development of ocean data assimilation capability via establishment of a national activity; those include:

1. The April 1997 meeting on ocean general circulation models should be encouraged to discuss how data assimilation might assist forward model improvements. Would forward

modelers consider it important to have a national ocean data synthesis activity (ocean data assimilation center)?

2. Reports from this meeting and from the OGCM meeting scheduled for April 1997 will be made available for consideration by the NRC Ocean Studies Board’s Committee on Large-Scale Oceanographic Programs at a meeting in June 1997

3. The National Science Foundation is requested to take the lead in developing resources for a national ocean data assimilation capability.

VII REFERENCES

NASA, 1996. Ocean data Assimilation: Report of the Ocean Data Assimilation/Modeling Workshop, Irvine, CA 25-27 October 1994. 30 pp.

U.S. WOCE Office (1996). U.S. WOCE Synthesis Plan. U.S. WOCE Planning Report 16, 18 pp., U.S. WOCE Office, College Station, TX.

WOCE IPO (1997). WOCE Analysis, Interpretation, Modeling and Synthesis (AIMS): Implementation Plan. International WOCE Project Office, Southampton, U.K. (in press).

APPENDIX I - MEETING ATTENDEES

D. Behringer (NOAA/NCEP)
A. Bennett (OSU, representing WOCE)
M. Blackmon (NCAR)
F. Bryan (NCAR)
J. Carton (U. Maryland)
P. Chapman (US WOCE Office)
S. Doney (NCAR, representing JGOFS)
T. Ezer (Princeton)
L. Fu (NASA/JPL)
E. Harrison (NOAA/PMEL)
R. Lambert (NSF)
J. Lupton (NOAA/PMEL, representing RIDGE)
J. Marotzke (MIT, representing WOCE, CLIVAR Dec-Cen)
R. Molinari (NOAA/AOML, representing ACCE)
P. Niiler (SIO, representing CLIVAR Dec-Cen)
W. Nowlin (TAMU, Chair)
Z. Powell (UC Berkeley, representing GLOBEC)
W. Prell (Brown U., representing ODP)
M. Reeve (NSF)
M. Roman (Horn Point, representing COOP)
A. Rosati (NOAA/GFDL)
K. Thompson (Dalhousie U.)
L. Walstad (Horn Point, representing COOP).

APPENDIX II - AGENDA

Meeting on Community Needs for Ocean Data Synthesis via Models
Dallas, 20-21 February, 1997

Thursday 20 February

07:30 Continental breakfast items available

08:00 Welcome and background for meeting (Nowlin, Reeve)

08:20 Icebreaker: Everyone has the opportunity to talk for a few minutes on their personal or program needs for data assimilation and their aspirations for the meeting.

10:00 Break

10:15 Discussion of general elements of ocean data assimilation (Bennett)

11:00 Illustrative examples of data assimilation:

1. A shelf model (Thompson)
2. A quasi-operational ENSO prediction model (Behringer)
3. A basin/global ocean model (Marotzke)

12:00 Lunch

13:30 Continuation of examples as needed.

14:00 Presentation of statements on program needs for ocean data assimilation:

TOGA (Behringer)
COOP (Roman)
WOCE (Bennett, Marotzke)
ACCE (Molinari)
JGOFS (Doney)
GLOBEC (Powell)
Carbon dioxide (Caldeira)
RIDGE (Lupton)
ODP (Prell)
Other--interested parties, please let Nowlin know

17:00 Agency perspectives (Reeve, Lambert)

17.30 Close

Friday 21 February

07:30 Continental breakfast items available

08:00 Revisit points left from previous day's discussions

08:30 Discussion of common requirements, Assessment of present and future community infrastructure needs for ocean data assimilation

11:00 Summary of conclusions
Needed future activities
Establishment of writing teams as necessary

12.00 Lunch

Continued after lunch if needed

Close of meeting

APPENDIX III - PRESENTATIONS BY BENNETT, BEHRINGER, THOMPSON, MAROTZKE

1. General Elements of a Data Assimilation System - A. Bennett

Inverse modeling, data assimilation or, best of all, "ocean estimation" always has certain inputs and outputs. Sometimes the inputs are not explicitly stated, but can always be inferred from the estimation scheme. Not all outputs are always produced, but they could. A list of inputs and outputs is as follows

INPUTS

1) An ocean model, with a fast numerical solver. It may be necessary to integrate the model hundreds or even thousands of times, in order to estimate the ocean circulation and put error bounds on the estimate.

2) Data, with precise mathematical definitions of the measuring process. These definitions may be quite intricate for say, reciprocal-shooting tomography.

3) Prior probability distribution functions for errors in the model equations, and the data. These priors are often just means and covariances, sometimes with the tacit assumption that the distributions are Gaussian.

Items (1)-(3) comprise a formal and testable null hypothesis "H0" about the ocean, namely, that the real ocean fits the model and data to within the specified tolerances. The alternative hypothesis would be that the priors are wrong.

4) A cost or penalty function, that "estimates" the errors, leading to estimates of the ocean circulation. A reasonable choice is the estimator of maximum likelihood, which is least squares for Gaussian errors. The means and covariances included in H0 define the biases and weights in the least -squares expression.

5) An efficient algorithm for minimizing the estimator, for example, finding the weighted least-squares best-fit to the model and data.

OUTPUTS

6) Estimates or "analyses" or "maps" of all circulation fields. These occupy a lot of storage when the estimation is four-dimensional. Indeed, for efficient calculation, the computer should be able to hold them in core. The NRL CM-5E has 32 Gbytes of RAM, which is the right order of magnitude.

7) Maps of residuals in all the model equations, and tables of misfits to all the data. The model residuals may be adjustments to initial and boundary data, or to internally distributed forcing, or may represent corrections to wrong dynamics. An probable cause of model error is the difference between turbulent stresses and parameterization.

8) Test statistics for H0, in the form of the minimum value of the estimator. If the errors are Gaussian and H0 is correct, then the test statistic is the chi-squared variable with as exactly many degrees of freedom as there are data. Its expected value and variance are well known. This information also provides a rational stopping criterion in the search for the best fit.

If the test statistics for many assimilation experiments with the same H0 are unexpectedly large, then H0 should be rejected. We would then have learnt something new about ocean dynamics and ocean data. If the test statistics are of the expected size then the maps of circulation and residuals are reliable, for the present.

9) Posterior probability distributions for errors in the estimated circulation and residuals. These posteriors may be Gaussians, with means and covariances calculated by Monte Carlo simulation, for example.

10) Array assessment, in the form of conditioning information calculated during the fitting process. The data may be redundant on the scales imposed by H0 and the model, in which case conditioning may be poor if the data are assumed accurate.

11) Model improvements, deduced from the patterns of circulations and residuals.

All these items are included in an estimation of coupled tropical Pacific ocean-atmosphere circulation, with a modified Cane-Zebiak model and a year of multivariate Tropical Atmosphere Ocean (TAO) data. It has always been obvious that C&Z -like models neglect eddy heat fluxes and atmosphere-to-ocean heat exchange. However, an examination as in (11) indicates that both must be included in order for the model to track the SST data.

2. Ocean Data Assimilation at NCEP – D. Behringer

For a number of years at NCEP, we have used an ocean analysis system based on GFDL's MOM GCM and the 3D variational scheme of Derber and Rosati (1989) to initialize coupled ENSO forecasts. We are now testing a new ocean data assimilation system which extends the Derber and Rosati (1989) scheme. The new system assimilates satellite observations of sea surface height, represented as deviations from a long term mean, as well as the traditional surface temperature and temperature profile data. The first guess field is the 3D
model temperature field and it is only this field which is updated or corrected. The correction is obtained by simultaneously minimizing the differences between:

1) the corrected temperature and the current model solution,

2) the corrected temperature and the observations of temperature, and

3) an estimate of dynamic height based on the corrected temperature and the satellite estimate of surface height, again, represented as deviations from a long term mean.

The measure for each of these differences is weighted by an appropriate error covariance matrix. In the original Derber and Rosati (1989) scheme errors in the observed data are assumed to be uncorrelated and this assumption is retained in the extended scheme. With respect to the model or first guess error correlations, the Derber and Rosati (1989) scheme ignores vertical error correlations. It also assumes that horizontal error correlations are the same at each model level and represents them by a Gaussian function. This approach works well enough as long as the corrections in each model level are based solely on direct comparisons with temperature measurements at that level. In the extended scheme, the satellite sea surface height data provide an integral constraint on the temperature corrections across the model levels and consequently the vertical structure of the model errors can no longer be ignored. The vertical structure of the model error covariance will determine the vertical distribution of the temperature corrections which are needed to minimize the difference between dynamic height and observed sea surface height.

Past experience in assimilating only temperature profile data has shown that model-data differences are largest in the thermocline. We have therefore chosen to approximate the vertical structure of the model error by making the error variance at each model level proportional to the local vertical gradient of temperature and representing the vertical correlation by a Gaussian shape. This simple representation ensures that temperature corrections implied by differences between model and satellite estimates of sea surface height are vertically correlated and are concentrated in the thermocline.

We have applied this system to the tropical Pacific Ocean for the TOPEX period. If we compare two experiments, one assimilating only temperature profile data from XBTs and TAO moorings and the other assimilating all of these data plus sea surface height deviations from TOPEX, we find that the low frequency sea surface height within the tropics is comparably represented in both experiments. The rms difference between model SSH and the monthly time-series for 8 independent tide gauges within 5 degrees of the equator is 2.65 cm for the experiment assimilating only profile data, the rms difference is reduced to 2.45 cm when the TOPEX data is also included. When the experiments are compared to 29 tide gauges throughout the tropics the rms differences are 3.3 cm without TOPEX and 2.8 cm with TOPEX.

3. Assimilation of data into a coastal model at Dalhousie – K. Thompson

At Dalhousie University, Thompson and his co-workers Jinyu Sheng, Michael Dowd, Josko Bobanovic and Mark Buehner, are assimilating data into dynamical models of shelf circulation. A summary was given of the development and use of a limited-area operational model for nowcasting and forecasting circulation on the outer Scotian Shelf. Data are assimilated using the so-called adjoint method with the controls taken as the open
boundary conditions and the baroclinic component of flow. The main conclusion of this work is that simple, limited-area models of shelf circulation can synthesize a wide range of data types and provide better estimates of the circulation than purely statistical schemes, such as optimal interpolation.

A prototype three-dimensional operational model is now running at Dalhousie. The model has a diagnosed baroclinic flow as a background state. It is driven by observed and forecast winds (from the Canadian Atmospheric Environment Service) and assimilates coastal sea-level in order to define flows across model open boundaries. Using an independent set of moored current measurements and drifter trajectories collected in February 1996, it was shown that the assimilative model has significant predictive skill in nowcasting and forecasting the flow several days into the future.

Ongoing work at Dalhousie includes the development of assimilation schemes for shelf circulation models which include a density field that evolves with the flow. Results based on an approximate adjoint model look promising and suggest that it may be possible to develop an operational model for shelf circulation that can assimilate not only coastal sea-level, moored current meter and drifter data, but also satellite images and spot hydrographic measurements.

4. Data assimilation into regional and global models – J. Marotzke

Jochem Marotzke emphasized that fitting general circulation models to observations has already produced results unachievable by other means. The MIT group has successfully used adjoints to GCMs, and has employed a new software tool (Giering and Kaminski, 1997) to generate the adjoint to the MIT GCM (Marotzke et al., 1997). This approach increases the flexibility of the adjoint approach enormously.

Marotzke showed two examples of the use of adjoint GCMs. The first, (Lee and Marotzke, 1997) used the GFDL ocean GCM and demonstrated that by fitting the model to hydrographic and surface flux data, the detrimental impact of incorrect open boundary conditions in an Indian Ocean GCM could be significantly reduced; a model-data combination is necessary to estimate rather than simply prescribe open boundary conditions. Moreover, the results indicated that the Indian Ocean meridional transports are predominantly wind-driven; this result could not be obtained using models simpler than a GCM with the observations.

The second example (D. Stammer, 1997, pers. comm.) used the new MIT adjoint GCM to fit a 2-degree global configuration to the altimeter data for 1993. The model runs are not yet complete and results are very preliminary, but the approach demonstrates, on a much reduced scale, that global ocean state estimation on a roughly weekly basis is possible. Two fields from the model solution were shown, one demonstrating global coverage of estimated currents at all depths, the other estimated modifications of surface wind stress. Any particular solution feature must be closely examined to determine its physical robustness, but the fundamental possibility has been demonstrated of relating different global data sets to each other through the use of a GCM, thereby testing every component of the estimation system.

References:

Giering, R., and T. Kaminski, 1997: Recipes for adjoint code construction. Submitted to ACM Trans. Math. Software.

Lee, T., and J. Marotzke, 1997: Inferring meridional mass and heat transports in the Indian Ocean by combining a general circulation model with climatological data. Journal of Geophysical Research, in press.

Marotzke, J., Q. K. Zhang, R. Giering, D. Stammer, C. N. Hill, and T. Lee, 1997: The linearization and adjoint of the MIT general circulation model. Manuscript in preparation.